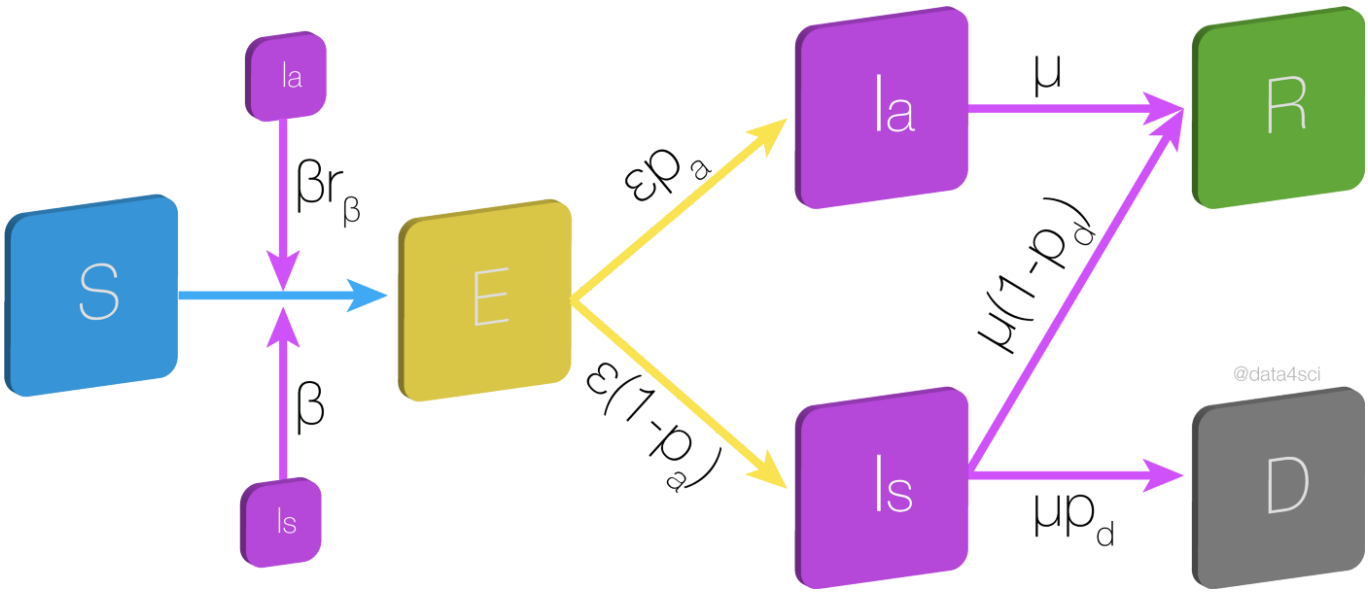
Anyone can publish on Medium per our Policies, but we don't fact-check every story.
For more info about the coronavirus, see cdc.gov.



Epidemic Modeling 102: All CoVID-19 models are wrong, but some are useful

Bruno Gonçalves Follow
Apr 5 · 12 min read ★

This is the second post of the "Epidemic Modeling" series. We will be building up on our discussion from the first post, "Epidemic Modeling 101: Or why your CoVID-19 exponential fits are wrong", so you might want to start reading there. You can find the notebooks I wrote to implement the models and generate the figures over at the GitHub repository I made specifically for this series:

Repository to accompany the blogpost: Epidemic Modeling 101: Or why your CoVID19 exponential fits are wrong

github.com

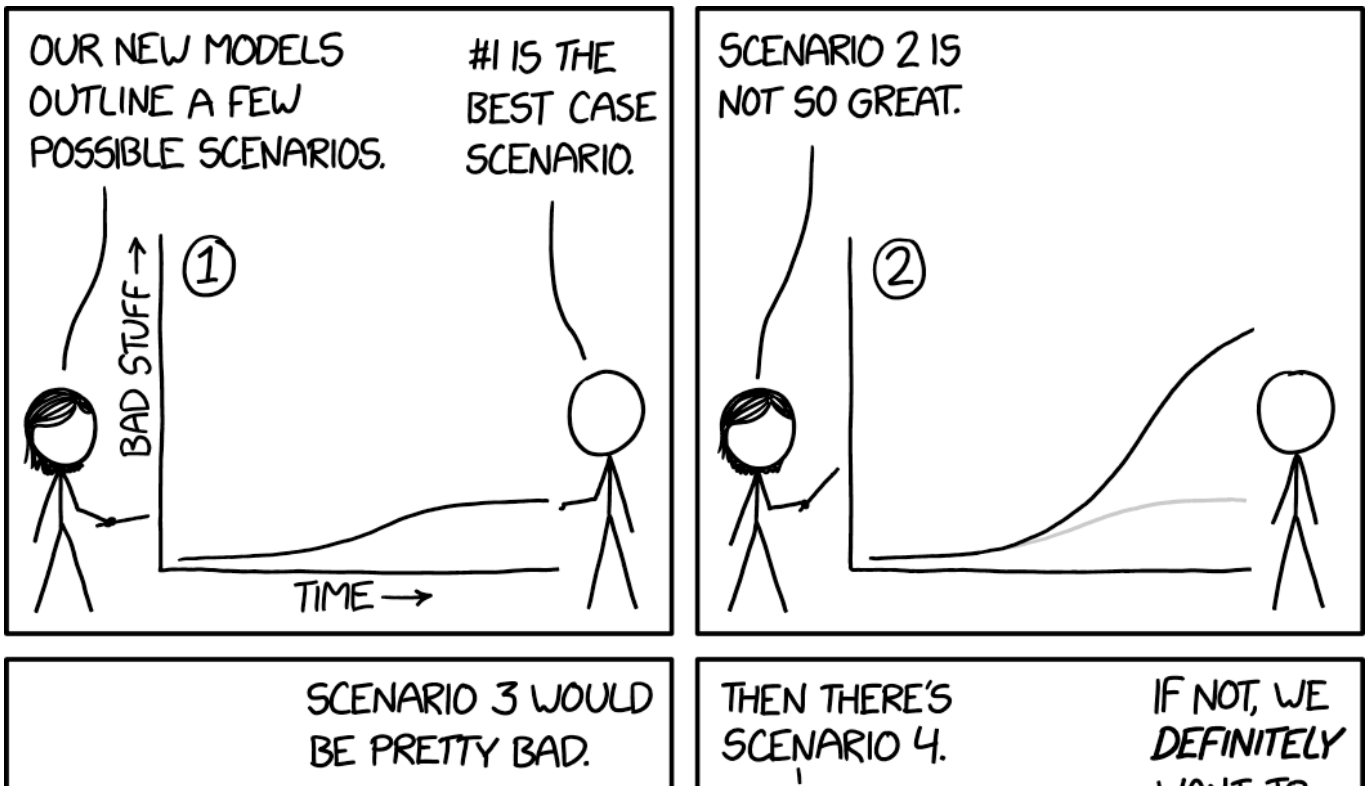
What follows is my personal perspective, as an individual with some real world experience in epidemic modeling during previous pandemics and shouldn't reflect on any group or institution I might be affiliated with.

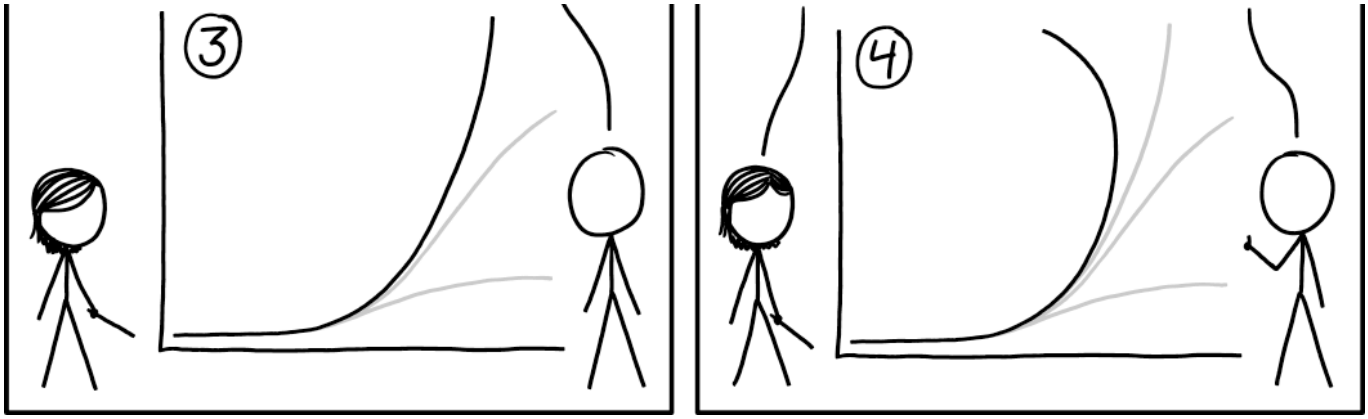
So, without further ado…

Models vs the real world

As George E. P. Box, a statistician, famously said "all models are wrong, but some are useful". This is perhaps never more true than during a crisis. Information is limited, often wrong, but decisions must be made and implemented based on what is known at the time.

It is also during an ongoing crisis that models play their most fundamental role, that of allowing us to explore scenarios and work through the consequences of our decisions:



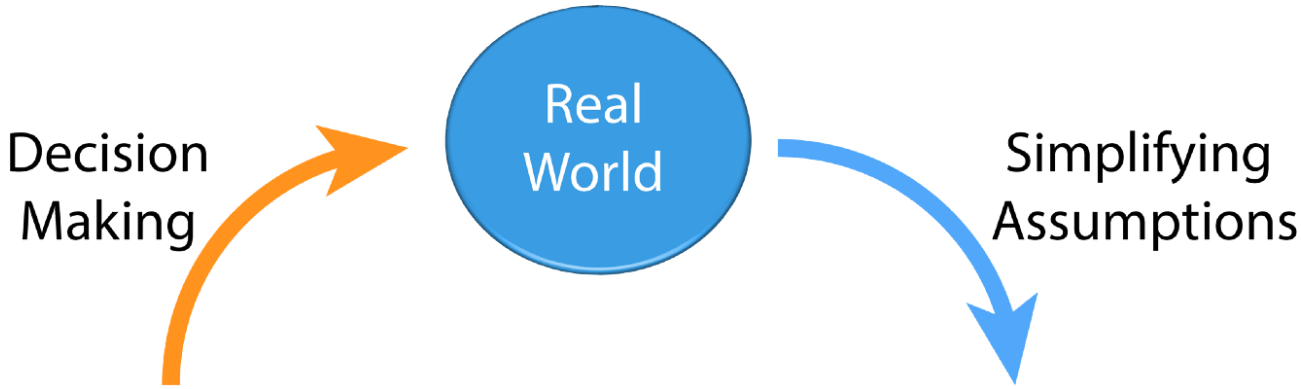


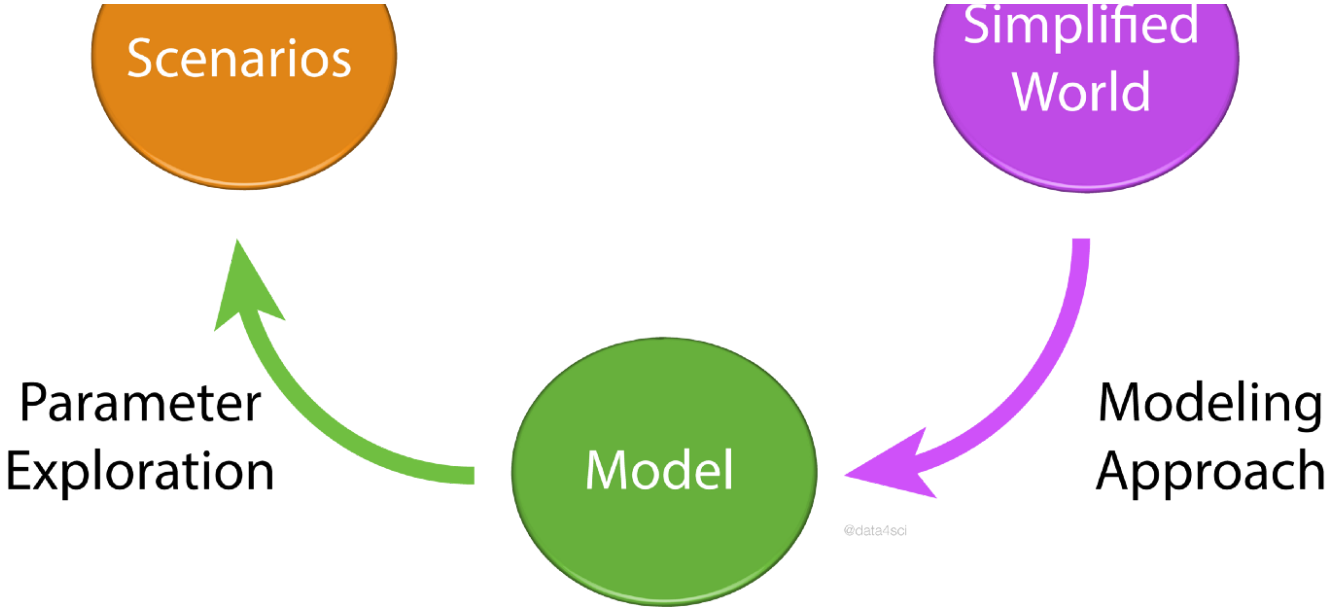
XKCD: "Remember, models aren't for telling you facts, they're for exploring dynamics. This model apparently explores time travel"

However, care must be taken to avoid mistaking the model for the reality. After all, "the map is not the territory". The development of a model, regardless of the domain of application, typically follows a common pattern:

A simplified version of the world is created, to which a specific modeling approach can be applied, resulting in a working model. The simplifications made can be due to a variety of factors such as the lack of specific data, excessive complexity, intractability, among others. The modeling approach chosen is both influenced by and helps drive the assumptions that are made, often resulting in the stereotypical overabundance of Physicists concerned about Spherical Cows or trying to apply Ising Spins to every possible problem.

Once a working model is obtained, we can use it to explore scenarios, the consequences of specific decisions, etc. Finally, it is by studying the scenarios that are outputted by our models that decisions are taken in the Real World. Graphically, we have:



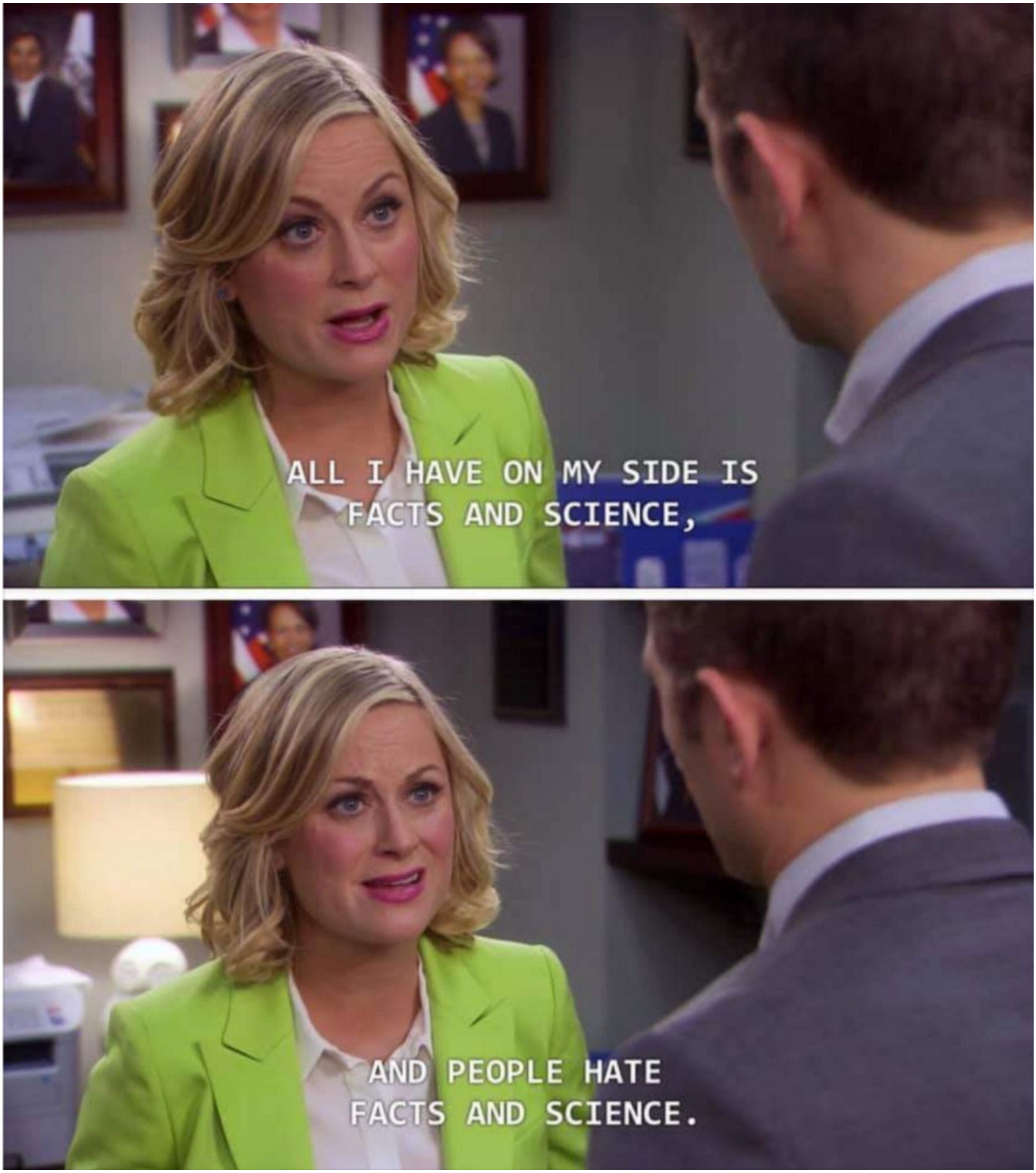


Naturally, this is a simplified and schematic view (and a model in and of itself) to help illustrate the various points at which our modeling efforts can go awry, leading the results of our models to differ from what we actually observe in the real world.

While in many cases, mismatches between the model and reality can be traced back to errors made during the process, they can also be due to the fact that our model was successful and it resulted in appropriate measures being taken to prevent the undesirable scenarios it predicted. This is specially true in the case of highly visible models that are used to guide government decisions and interventions such as in the case of an ongoing pandemic like the one we're living through now:

> *"The most important function of epidemiological models is as a simulation, a way to see our potential futures ahead of time, and how that interacts with the choices we make today. With COVID-19 models, we have one simple, urgent goal: to ignore all the optimistic branches and that thick trunk in the middle representing the most likely outcomes. Instead, we need to focus on the branches representing the worst outcomes, and prune them with all our might. Social isolation reduces transmission, and slows the spread of the disease. In doing so, it chops off branches that represent some of the worst futures. Contact tracing catches people before they infect others, pruning more branches that represent unchecked catastrophes." — Zeynep Tufecki, The Atlantic*

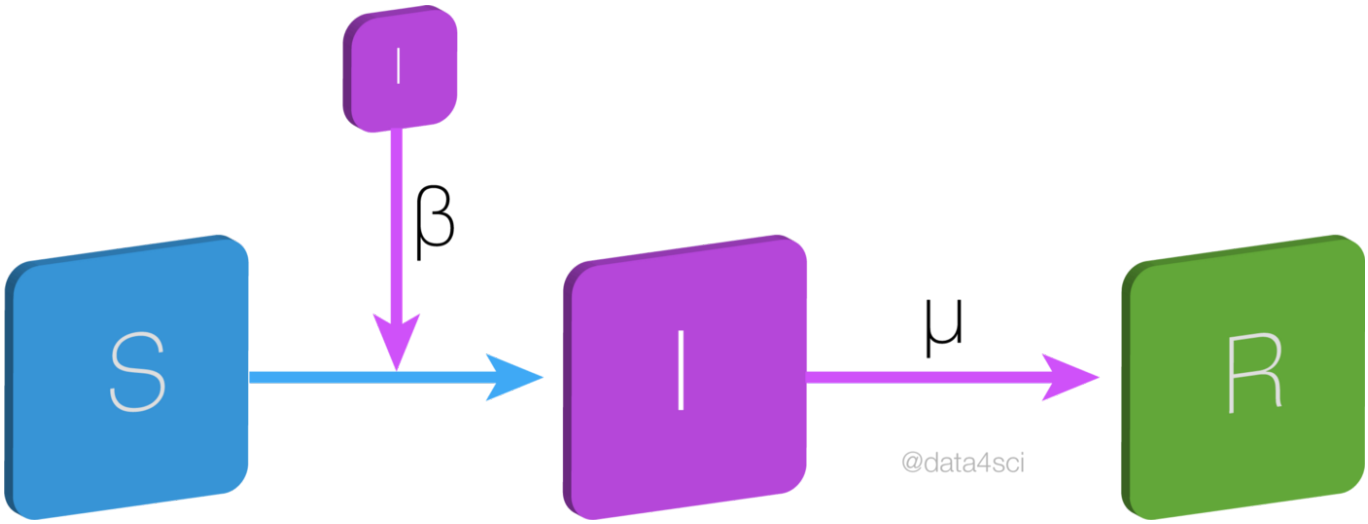
It is this kind of misunderstanding that leads to public distrust in scientific models in particular and Science in general.



My hope is that this (and many other posts out there) can help the general public to understand the underlying assumptions, power, and limitations of scientific models and how they can be put to good use.

Susceptible-Infectious-Recovered (SIR) Model

Now that we have established both the advantages and limitations of using models to understand the world, we can start exploring how to improve the simple models we introduced in the previous post.



SIR Model

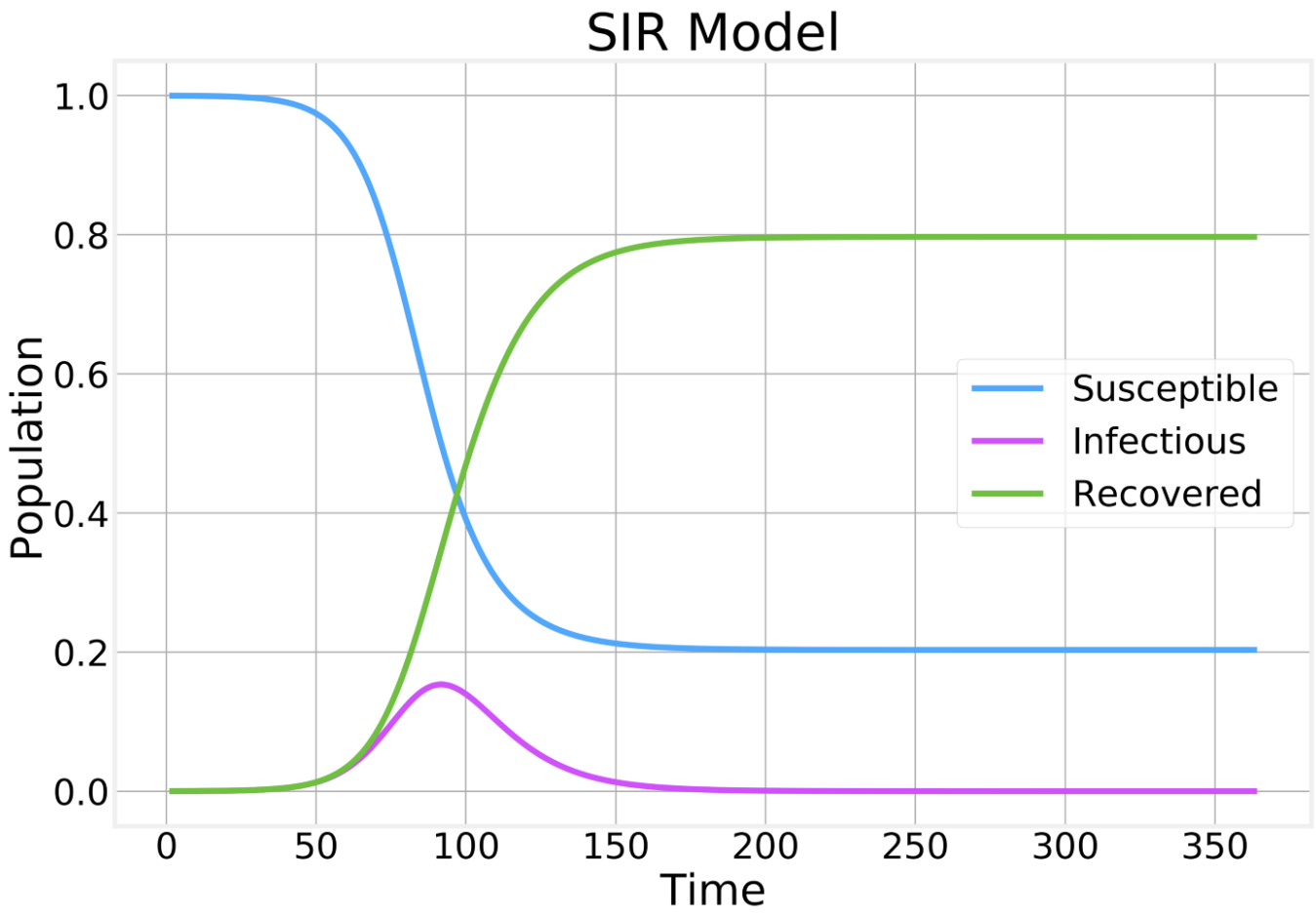
The SIR model is one of the simplest and best known epidemic models. Its popularity is due, in no small part, to its ability to establish a perfect balance between simplicity and usefulness. It is still relatively amenable to mathematical and analytical exploration while at the same time it is able to capture the fundamental features of the epidemic process: healthy (*Susceptible*) people become infected when coming in contact with *Infectious* individuals only to eventually *Recover* after a certain period of time. The process is illustrated schematically in the figure at the top of this section.

This model can be written mathematically using a simple set of partial differential equations:

$$\begin{aligned}
\frac{\partial}{\partial t} S_t &= -\beta S_t \frac{I_t}{N} \\
\frac{\partial}{\partial t} I_t &= +\beta S_t \frac{I_t}{N} - \mu I_t \\
\frac{\partial}{\partial t} R_t &= +\mu I_t
\end{aligned}$$

Susceptible-Infectious-Recovered model

Which can be numerically integrated to obtain the values of each compartment as a function of time, just as done in the previous blog post:



Fraction of the population in each compartment as a function of time

While this kind of equations can be useful to explore analytical results for simple models like the SIR model, they quickly become unwieldy for more complex models. However, it is easy to note how they have a one-to-one correspondence with the illustration above:

- Interactions correspond to terms involving two compartments and the total number of individuals in the population:

$$\beta S_t \frac{I_t}{N}$$

Interaction term

- While spontaneous transitions correspond to terms involving just a single compartment:

$$\mu I$$

Spontaneous term

- The sign of each term is determined by whether the equation we are considering corresponds to the "source" or "target" compartments. Notably, "agent" compartments are not affected unless they are also "targets".

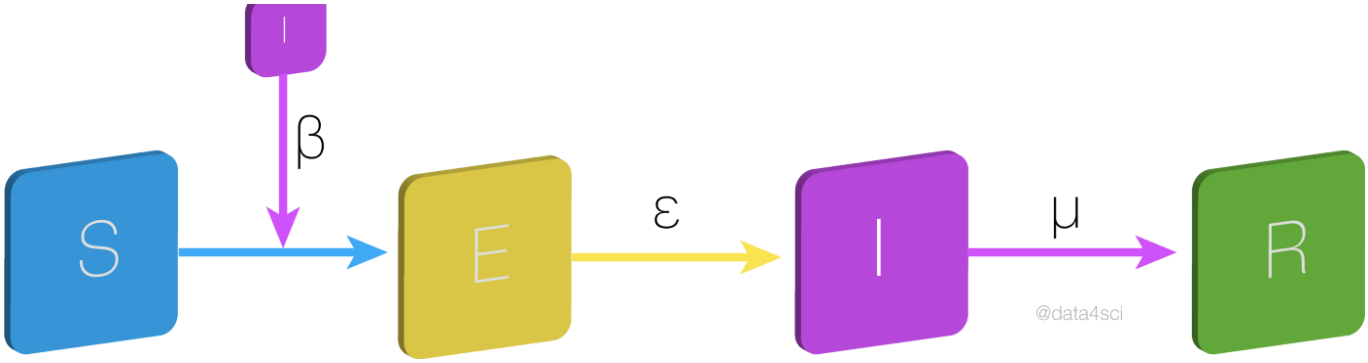
This one-to-one correspondence between transitions and terms allows us to simply "draw up" arbitrarily complex models that can be trivially implemented using generic code (like the one in EpiModel.py) without having to write out and debug all the rules "by hand".

In the rest of the discussion we will focus on discussing the assumptions and details of the various models while avoiding as much as possible the use of complex mathematical expressions.

Incubation Period

One of the main limitations of the SIR model is the fact that the infection develops instantaneously without any incubation period what so ever. You'll recall from recent news that this is not a very realistic scenario and the incubation or latent period is one of the most important factors that must be understood in order to contain an epidemic: For how long must a suspected case be kept under watch until we can be certain that the person will not become infectious?

We can address this limitation by adding one extra step (compartment) to our epidemic model: The *Exposed* (or *Latent*) compartment. When a *Susceptible* person comes in contact with an infectious one s/he moves to the *Exposed* from which s/he transitions to the *Infectious* compartment at a fixed rate ε. While in the *Exposed* compartment the person is said to be "incubating" the disease, possibly even starting to develop symptoms, but is not yet able to infect other individuals. The resulting model is known as the *Susceptible-Exposed-Infectious-Recovered* (SEIR) model:



SEIR Model

Here we have 4 distinct compartments connected by one interacting transition and two spontaneous ones:

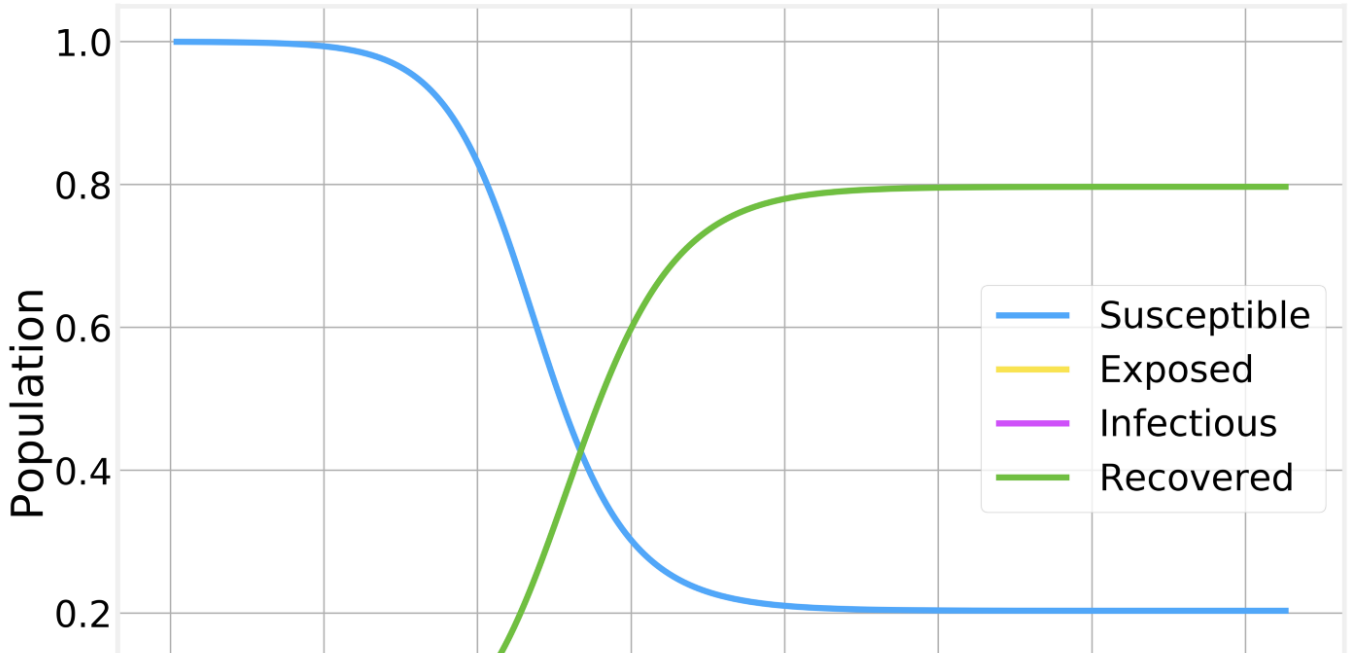
$$S + I \xrightarrow{\beta} E$$

$$E \xrightarrow{\epsilon} I$$

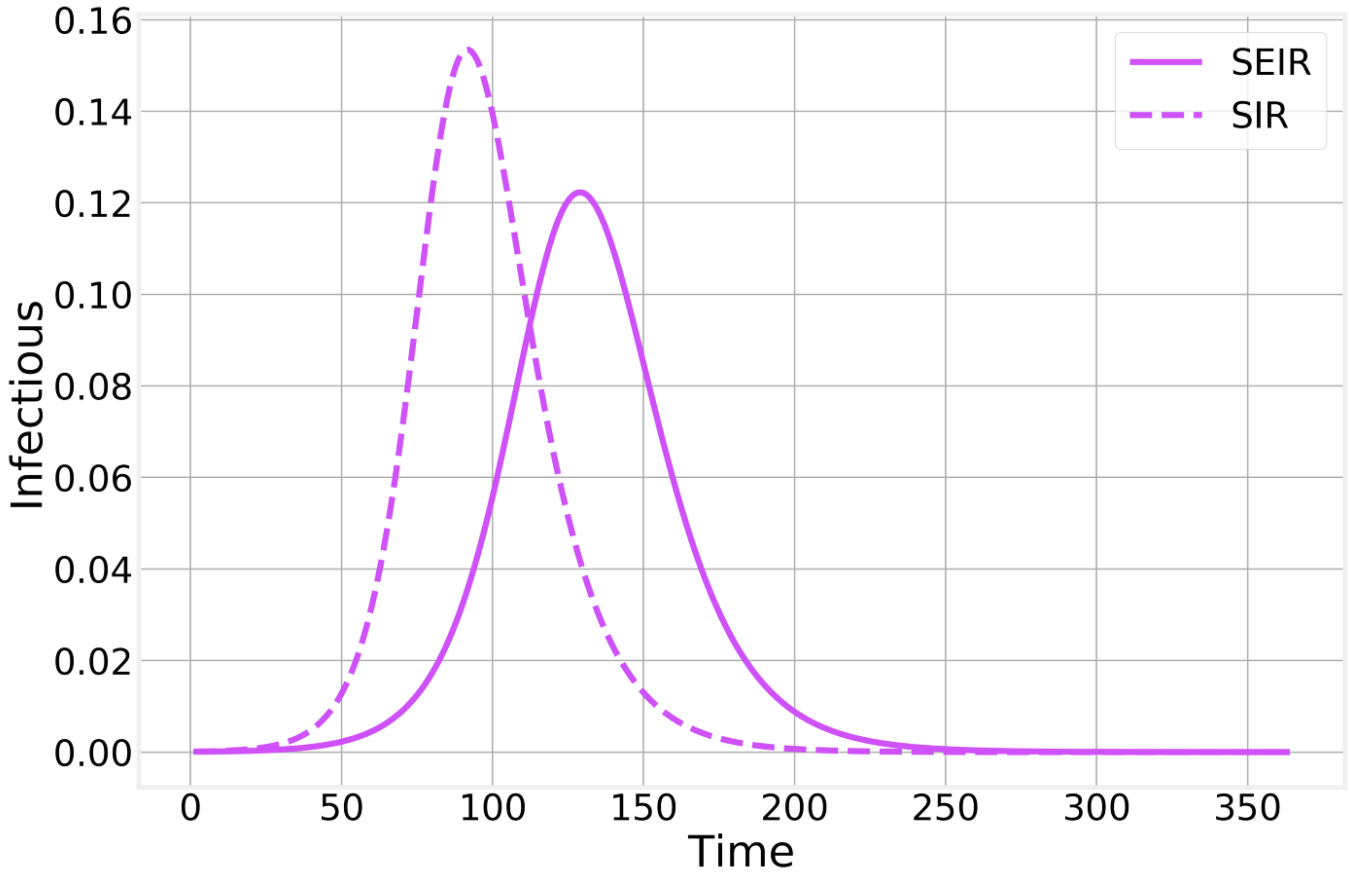
$$I \xrightarrow{\mu} R$$

SEIR Transitions

And the evolution of the various compartments is simply:



Here we highlight that the addition of the extra compartment didn't change the total number of people who become infected, but it does have a strong impact on the temporal evolution of the epidemic, significantly delaying and widening the peak of infectious cases. It effectively "flattens" the curve:



Epidemic peak comparison between the SIR and SEIR models.

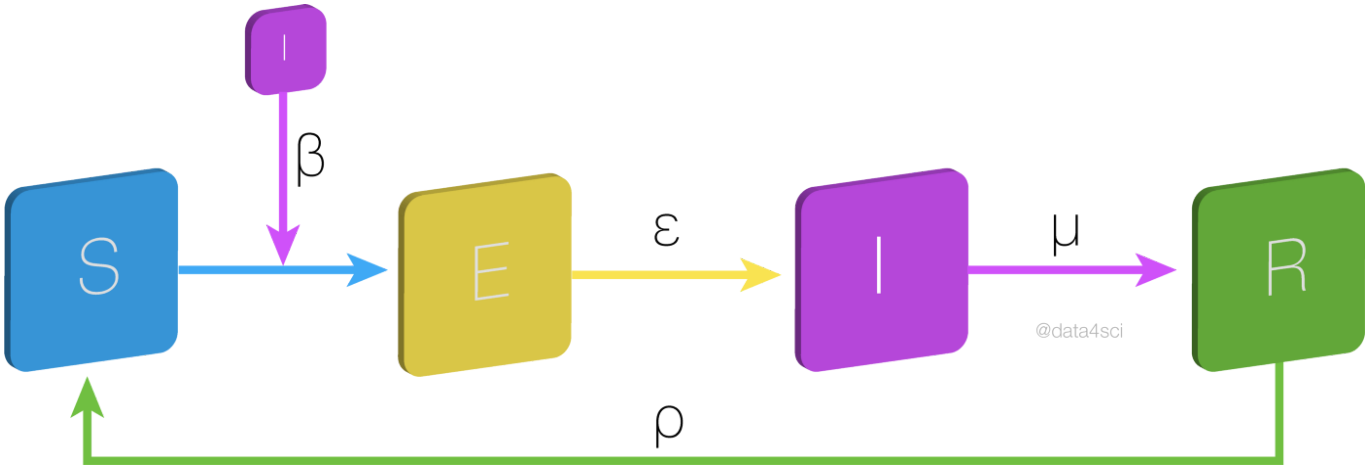
It should be clear how this has a direct impact on the likelihood of the healthcare system being overwhelmed and the necessary duration of any quarantine measures imposed: **a lower peak reduces the stress in the healthcare system**, while **a longer duration implies that longer period of social distancing** is necessary.

Temporary Immunity

Another fundamental assumption underlying the SIR model is the idea that *Recovered* persons are permanently immune from the disease. While this is the case with many common diseases, there have been some reports of CoVID-19 patients being re-infected after recovery.

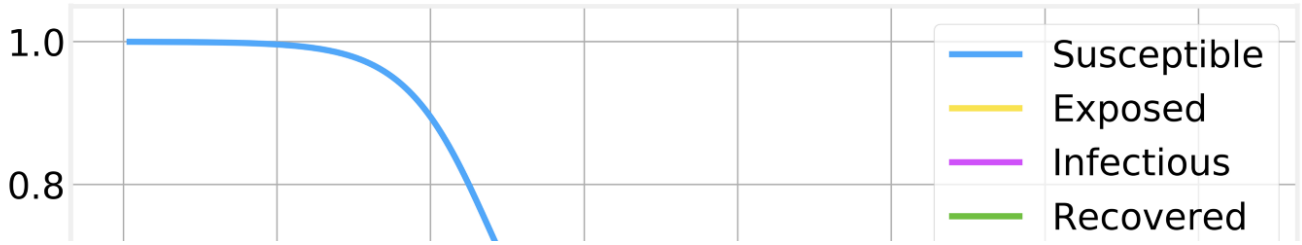
Re-infection in such a short period of time is unlikely (even temporary immunity typically lasts for a few months or years) and these cases are more likely to be due to faulty tests, but it is certainly a possibility that should be considered.

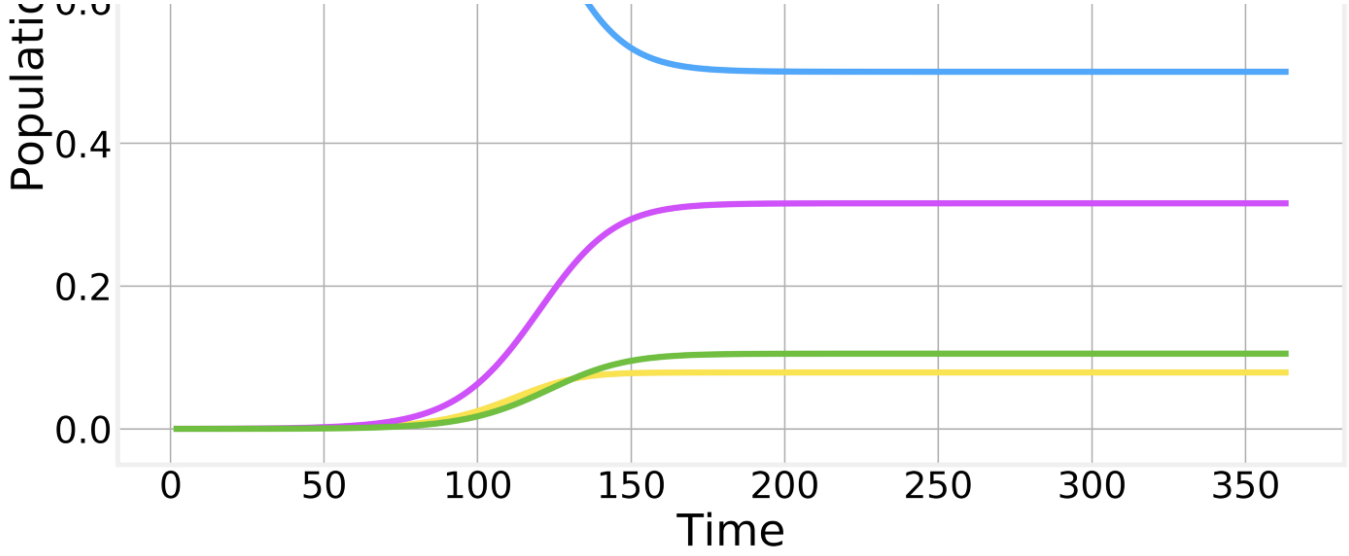
By simply adding a spontaneous transition from the *Recovered* compartment back to the *Susceptible* compartment, we obtain the **SEIRS** (can you guess what the letters stand for? 😀):



SEIRS Model

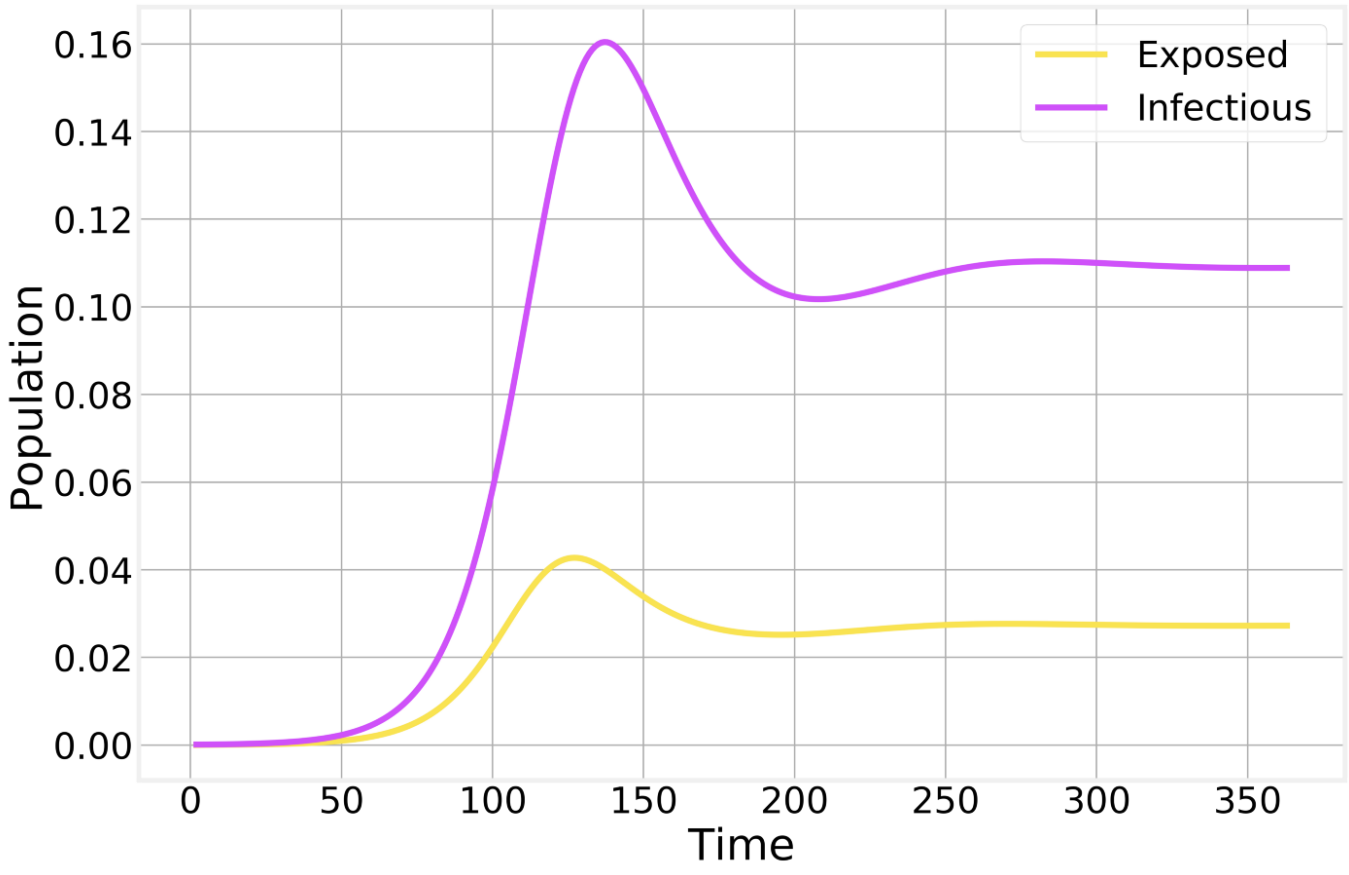
This seemingly innocuous addition to the model has a **very important effect**. By allowing *Recovered* individuals to once more become *Susceptible*, we replenish the group of people that can once again be infected. The end result is that the epidemic never burns itself out (its fuel is never exhausted) and the disease becomes endemic, with a constant fraction of the population remaining infected!





Endemic final state of the SEIRS model

The rate ρ at which immunity is lost has a determinant effect in the progress of the epidemic and the rise of endemicity. If ρ is sufficiently small (immunity is longer lasting) we can even have several epidemic peaks before the steady state of a fixed fraction of the population is reached.

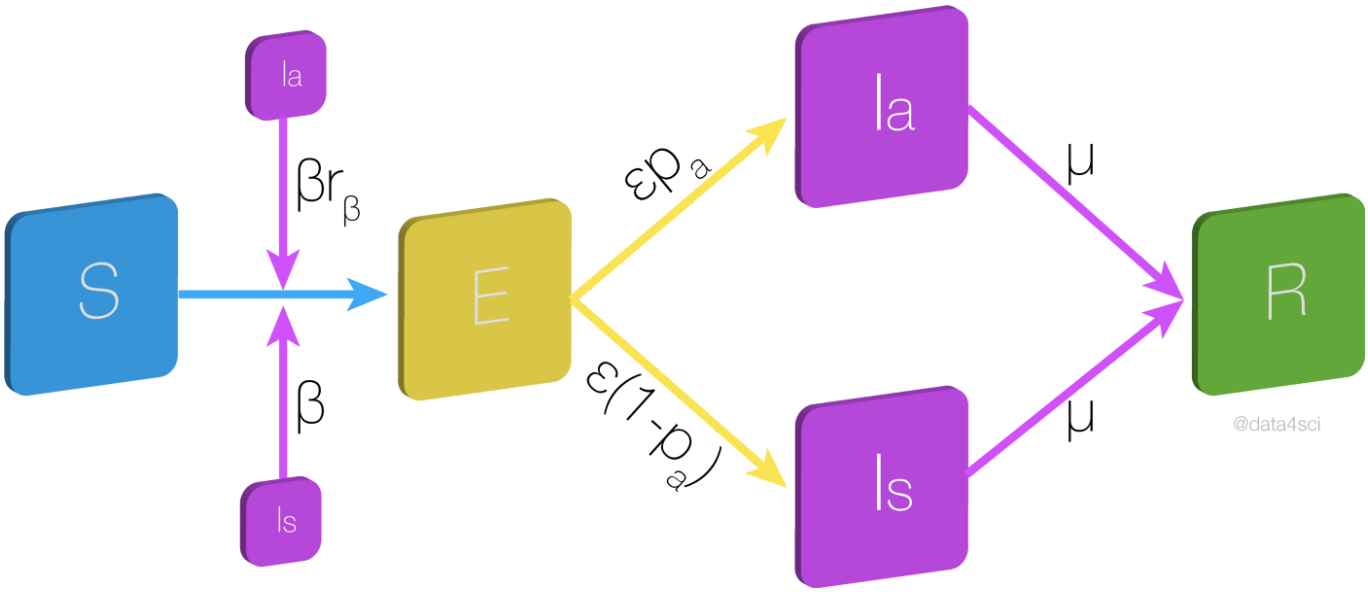


The appearance of the peak is due to the fact that the temporary immunity afforded by the disease is sufficiently long to allow the epidemic to follow most of its course before the number of *Susceptibles* starts to increase again, adding fuel to the fire.

Asymptomatic Cases

In many diseases, a significant fraction of infected individuals remain asymptomatic throughout the course of the disease. In the case of seasonal Influenza, this number is typically around 33%, while for CoVID19 the number is thought to be 40% or higher, thus significantly skewing the total number of cases.

Asymptomatic individuals are often less infectious than those displaying symptoms by some fraction r_β. We can model their effect by splitting the *Infectious* compartment in two: a Symptomatic, *Is*, and an Asymptomatic, *Ia*. A fraction p_a of all of those *Exposed* become asymptomatic while the remaining *(1-p_a)* develop symptoms. Our model is then:



As we now have two *Infectious* compartments we must also redo our R_o calculation. Fortunately, the modification is simple: since we have split the original *Infectious* compartment in two, our value of β is simply the weighted average of the original and the reduced β.

$$R_0 = \frac{\beta}{\mu}\left[p_a r_\beta + (1 - p_a)\right]$$

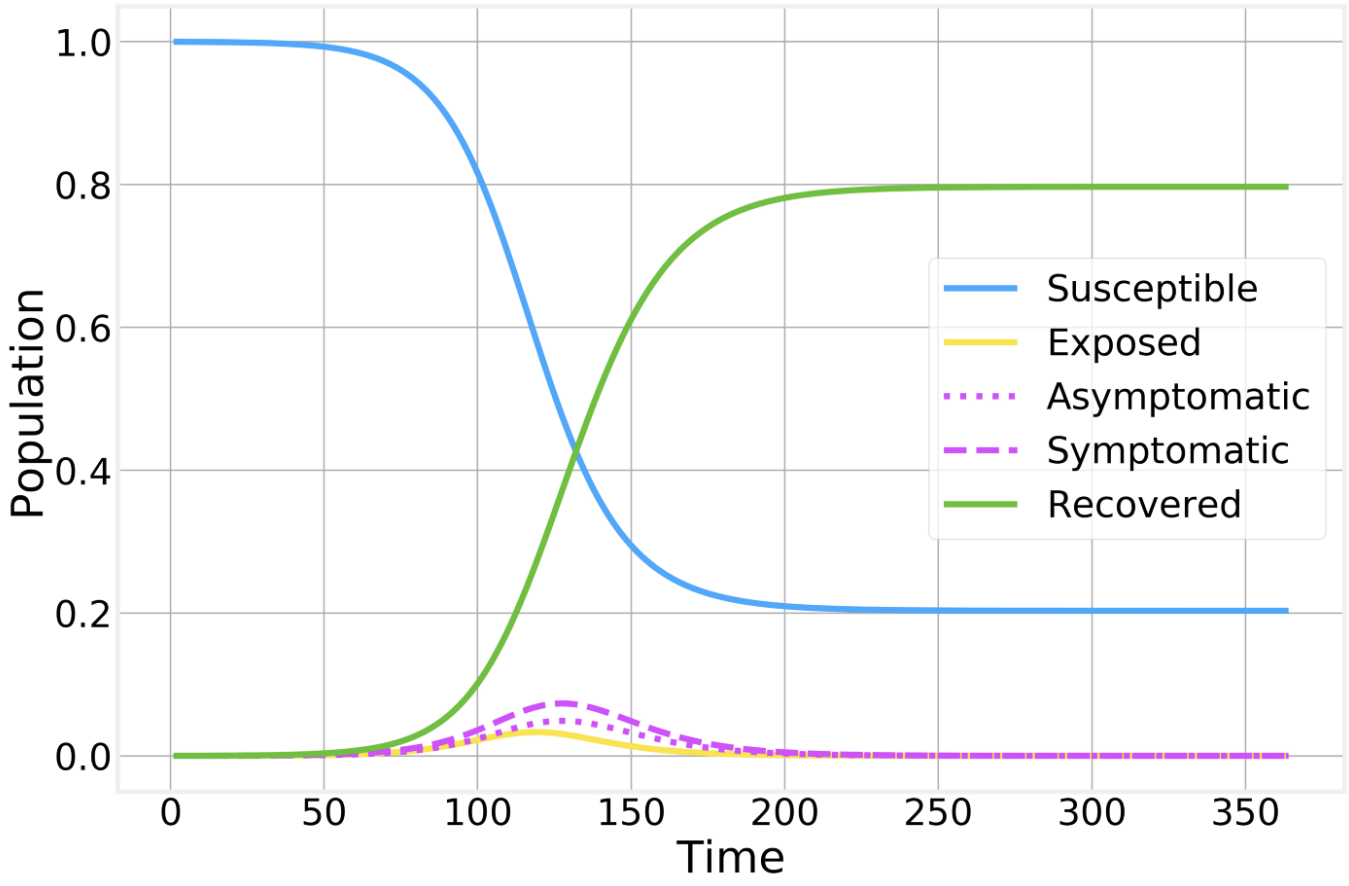
We can easily verify that if r_β is 1 we recover the original SIR value, while if r_β is 0 (the asymptomatic and completely non-infectious) we reduce the original R_o by a factor of (1-p_a) as we effectively have that much smaller *Infectious* population.

In order to maintain the same value of R_o we simply calculate the value of β as:

$$\beta = \frac{R_0 \mu}{p_a r_\beta + (1 - p_a)}$$

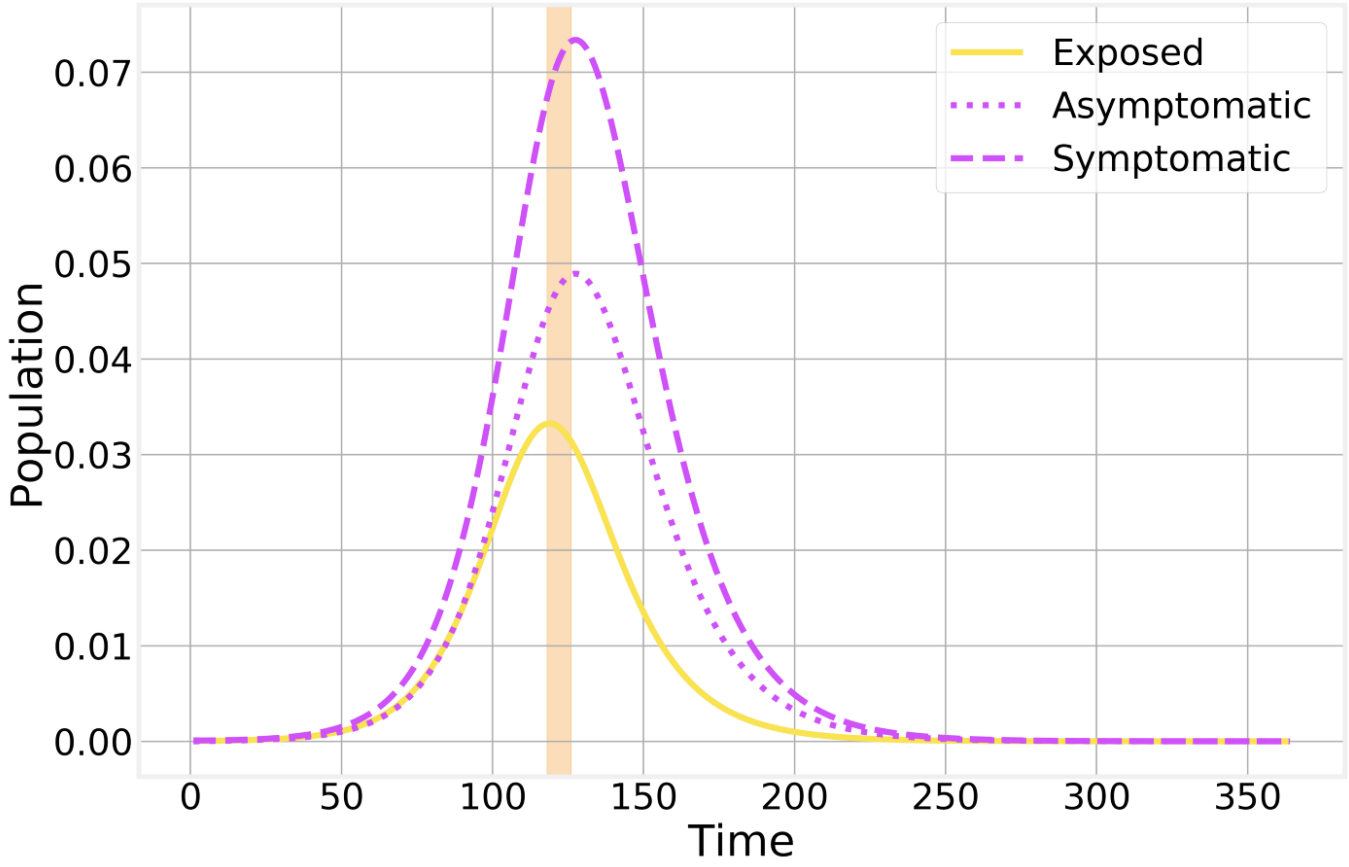
This approach makes it easier to compare the results from both models since they both have the same value of R_o.

As we add more and more compartments to our models, the smaller the population of each individual compartment becomes.



Compartmental structure of the Symptomatic/Asymptomatic model

We can easily verify that the value of R_o remains the same as before by looking at the *Recovered* and *Susceptible* curves at the end of the epidemic. On the other hand, we now have 3 distinct infected compartments, 2 of which are *Infectious* and peak at the same time and a few days after the *Exposed* population:

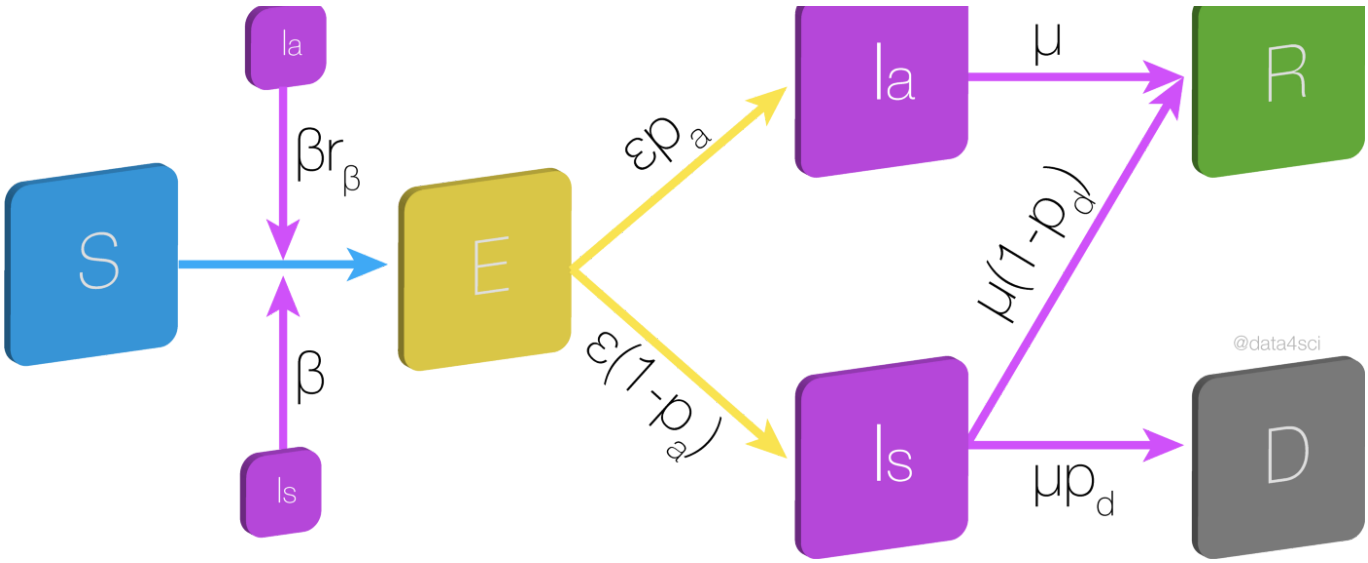


Peak comparison between the three infected compartments

Here we should note that we explicitly decided to keep the recovery rate, μ for both Symptomatic and Asymptomatic individuals. Had we chosen them to be different then the peaks would occur at different times and the expression for R_o would have to be revised even further.

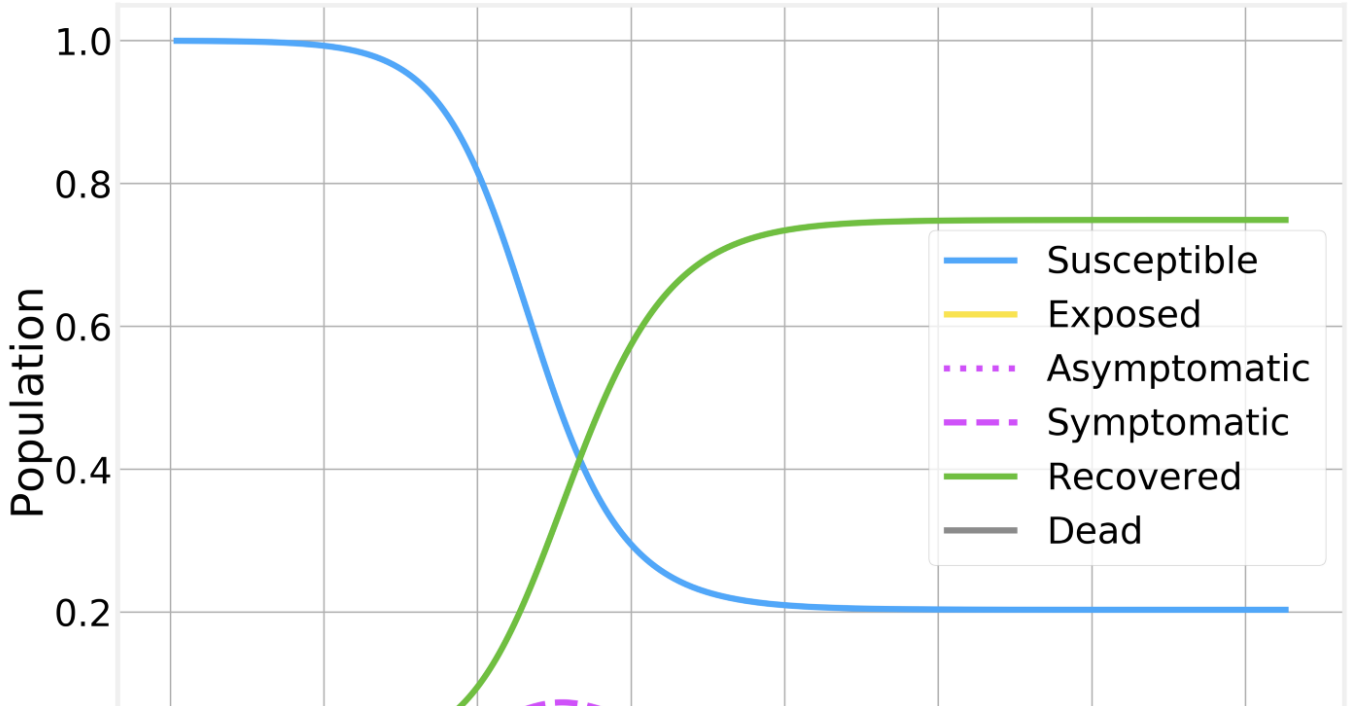
Mortality rate

Finally, we look at the effect of explicitly considering mortality. We assume that only symptomatic cases die from the disease or, similarly, that any asymptomatic individuals that do die from the disease are not counted as such. If we assume that a fraction *pd* of symptomatic cases end up dying, our model becomes:



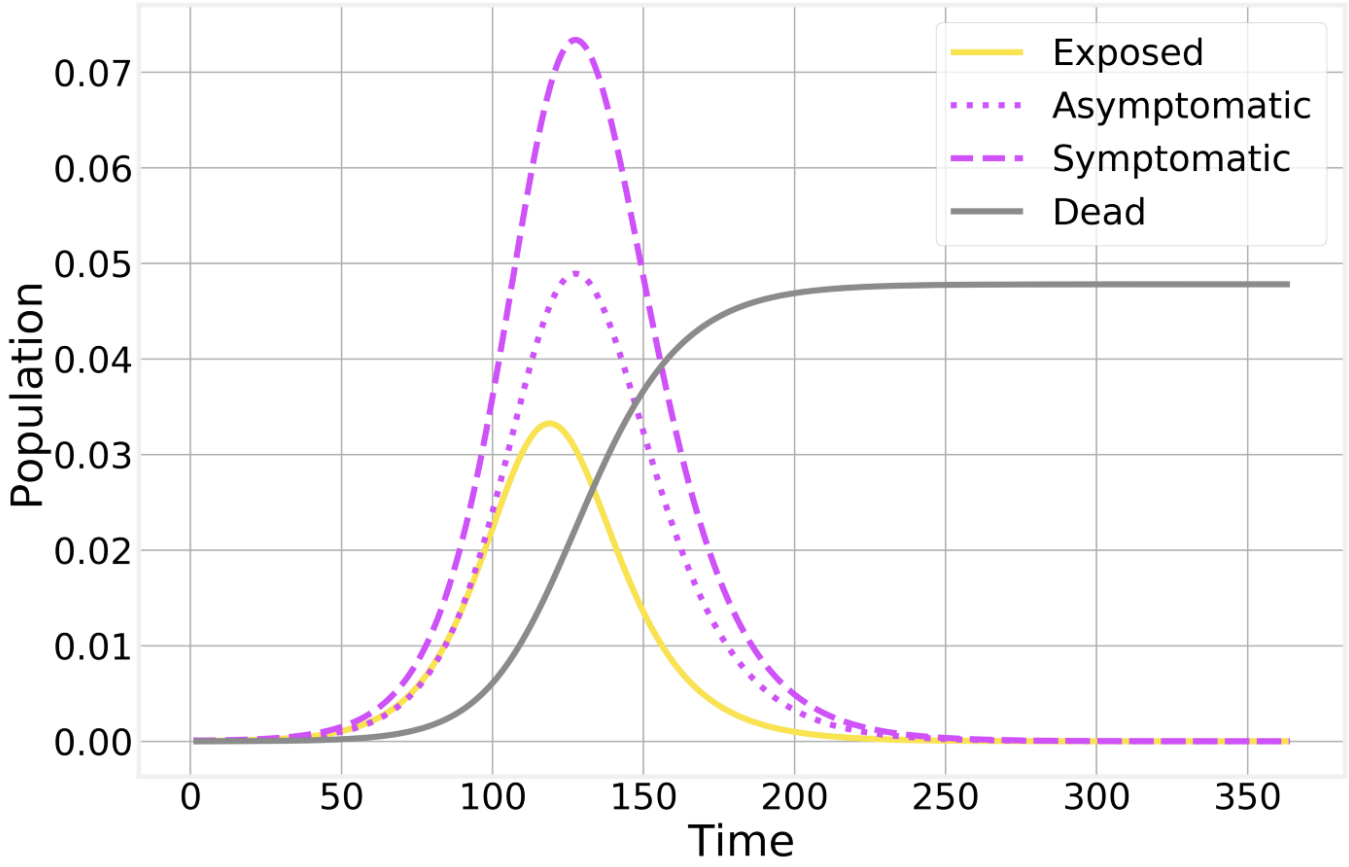
So we now have 6 compartments and a total of 7 transitions and 6 parameters, denoting how the more details we include the more complex the model becomes and the more parameters must be specified. In the early days of an epidemic most, if not all, of these parameters are partially or completely unknown. As the epidemic progresses, more and more information is gathered and more detailed models can be used. This constant refinement also helps improve the reliability of the scenarios we are able to analyze and the decisions made.

If we assume that *10%* of the symptomatic cases eventually die, we have:



It should be noted that ***10%* mortality rate is huge and unrealistic** for the kind of diseases we are considering. The reason we choose such a large number is **to make the effects obvious when plotting**.

By including the possibility of *Death*, the number of *Recovered* individuals is naturally reduced, despite the fact that none of the disease parameters have been changed. If we focus on just the relationship between the most significant compartments we have:



The total number of dead can be easily estimated. We know that for our set of parameters, 80% of the population eventually becomes infected. Of those, 60% are symptomatic and of those, 10% eventually die, so we expect that the total number of fatal cases to be 4.8% as shown in the plot above.

This value is significantly smaller that the actual mortality rate for the symptomatic cases. This is due to the fact that the number of recovered is inflated by the milder

Resources

Congratulations on making it through two loooong posts on Epidemic modeling. You already know a lot more about epidemic modeling than most mortals, but the journey has just begun.

All the code necessary to implement the models described above and to generate the figures used can be found in this posts GitHub repository:

DataForScience/Epidemiology101

Repository to accompany the blogpost: Epidemic Modeling 101: Or why your CoVID19 exponential fits are wrong...

github.com

If you enjoyed this post, you might also enjoy my weekly newsletter where I share the latest news and developments in Machine Learning and Data Science as well as any future blog posts I write.

First Name
Last Name
Email
Sign up
If you're ok with us sending you email updates, please tick the box. Unsubscribe whenever you want!

Covid 19 Epidemiology Python Modeling Programming

Stay up to date on coronavirus (Covid-19)

Follow the Medium Coronavirus Blog or sign up for the newsletter to read expert-backed coronavirus stories from Medium and across the web, such as:

- Why coronavirus is making young Americans really sick.
- Everyone's anxious. Therapists are slammed. Here's how mental health care is keeping up.
- These charts are a helpful resource for tracking the Covid-19 response.
- Coronavirus and the environment.